# COLUMBIA RIVER STOCK IDENTIFICATION STUDY:

# VALIDATION OF GENETIC METHOD

by

George B. Milner, David J. Teel, Fred M. Utter,

and

Curtis L. Burley[1/]

Final Report of Research (FY80)
Financed by
Bonneville Power Administration
(Contract DE-A179-80BP18488)

and

Coastal Zone and Estuarine Studies Division
Northwest and Alaska Fisheries Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
2725 Montlake Boulevard East
Seattle, Washington 98112

June 1981

1 / U.S. Fish and Wildlife Service, 9317 Highway 99, Suite 1, Vancouver, Washington 98665.

CONTENTS


Page

ABSTRACT ..... i

INTRODUCTION ..... 1

APPROACH ..... 2

METHODS ..... 3

Concept of the Genetic Identification Method ..... 3

Electrophoresis ..... 4

Maximum Likelihood Analysis ..... 4

Baseline Data ..... 5

Blind Samples A, B, C, and D ..... 7

Relationship of Calculated to Observed Standard Deviation ..... 7

Accuracy and Precision: An Effect of Potential Contributors ..... 7

Lower Columbia Gill-Net Fishery: A Field Test ..... 8

Time and Cost Study ..... 9

RESULTS ..... 11

Blind Samples A, B, C, and D ..... 12

Relationship of Calculated to Actual Standard Deviation..... 19

Accuracy and Precision: An Effect of Potential Contributors..... 26

Lower Columbia Spring Chinook Gill-Net Fishery..... 26

Time and Cost Estimates for Hypothetical Application..... 27

DISCUSSION ..... 30

LITERATURE CITED ..... 35

APPENDIX A

APPENDIX B

ABSTRACT

The reliability of a method for obtaining maximum likelihood estimates of component stocks in mixed populations of salmonids through the frequency of genetic variants in a mixed population and in potentially contributing stocks was tested in 1980. A data base of 10 polymorphic loci from 14 hatchery stocks of spring chinook salmon of the Columbia River was used to estimate proportions of these stocks in four different "blind" mixtures whose true composition was only revealed subsequent to obtaining estimates. The averaged differences of estimated and actual values of the 14 stocks in the four mixtures ranged from 0.9 to 7.4%; the precision of the estimates (measured by calculated standard deviations) ranged from 2.1 to 9.1%. Both accuracy and precision tended to improve when geographic groupings rather than individual stocks were considered, and dropped to a mean difference of 2.3% and a standard deviation of 1% when only two groups (stocks above or below Bonneville Dam) were considered. The accuracy and precision of these blind tests have validated the genetic method as a valuable means for identifying components of stock mixtures.

Properties of the genetic method were further examined by simulation studies using the pooled data of the four blind tests as a mixed fishery. Replicated tests with sample sizes between 100 and 1,000 indicated that actual standard deviations on estimated contributions were consistently lower than calculated standard deviations; this difference diminished as sample size increased. Both the accuracy and precision of estimates of a mixture involving two populations were greatest when only those two populations were included in the data base; the best estimates from a data base of all 14 populations were for that stock that was genetically most

distinct from the remainder of the group.

Estimates of 87% above and 13% below Bonneville Dam were made on a sample of 123 fish collected in Astoria, Oregon, following a 24-hour winter fishery. The moderate precision of this estimate would have increased with a larger sample size and an a priori reduction in the presumed complexity of the fishery.

Costs of $10.29 per fish were estimated for sampling, analyzing, and estimating the stock composition of a mixed fishery sample of 500 fish if a turnaround time of 24 hours from landing of fish to obtaining estimates were required. These costs would be reduced if a longer time interval were permissible.

It is recommended that future applications of the method be preceded by simulation studies that will identify appropriate levels of sampling required for acceptable levels of accuracy and precision. Variables in such studies include the stocks involved, the loci used, and the genetic differentiation among stocks.

# INTRODUCTION

Management of mixed stock fisheries based on their stock compositions is necessary if depressed runs are to be protected while abundant runs are being exploited. The genetic method of stock identification has attributes that make it a uniquely valuable source of information for such management. The method uses naturally occurring genetic differences to distinguish stocks and is therefore equally applicable to the identification of both wild and hatchery stocks. These identifying genetic characters remain constant throughout the life of an individual; thus the method can be used at any stage of the life-cycle of anadromous fish. The method features rapid and inexpensive laboratory analysis. Appropriate applications of the method can therefore provide management information that is both timely and cost effective.

The genetic identification method was developed in a cooperative project between the U.S. Fish and Wildlife Service (USFWS) and the National Marine Fisheries Service (NMFS), and was funded by the Bonneville Power Administration (BPA). The project was started in 1976 based upon previously developed techniques for detecting simple genetic variation in protein molecules (see Utter et al. 1974 for a review of early uses of these techniques in the study of fishes). Two early goals of the project were: 1) to develop a statistical method to use genetic data for estimating the composition of mixed stock fisheries and 2) to compile a data base of genetic variation among chinook salmon and steelhead stocks for use in the analysis of mixed fisheries involving stocks of the Columbia River drainage. These goals have been achieved (Milner, Teel, and Utter 1980).

The continuing project goal is to provide a stock identification method that will be used by managers of Columbia River salmon and steelhead resources. For the method to be widely accepted and applied, users of this management tool must first have confidence that it is reliable. Therefore, one major objective of work in 1980 was to demonstrate in a controlled test that the method works.

Practicality is a second condition which must be met if the method is to be widely used. Therefore, the second major objective of work in 1980 was to apply the method to actual and simulated Columbia River fisheries to study the factors which determine the practicality of a specific application.

## APPROACH

A blind test was conducted in cooperation with the USFWS to demonstrate the reliability of the genetic identification method. Personnel from USFWS composed four test mixtures of juvenile fish from 14 Columbia and Snake River stocks of spring chinook salmon. These stocks are representative of both the geographic and genetic structures of spring chinook salmon in the Columbia River drainage. The proportional contributions of the stocks were different in each test mixture. The four test mixtures, then, simulate four different spring chinook fisheries in the lower Columbia River. NMFS electrophoretically analyzed the blind samples and estimated their stock compositions. Only then, did USFWS reveal their actual compositions for comparison with the estimates.

In addition to the blind sample study, the compositions of actual and simulated fisheries were estimated to study factors which affect the practicality of the method. As with any method of estimating the

composition of mixed fisheries--whether it uses coded-wire tags, scale differences, or genetic differences to distinguish stocks--three variables determine the practicality of a potential application:

1) The sample size required for the given degree of accuracy and level of confidence of the estimate.

2) The interval of time from the sampling of the fishery until the estimation of its composition.

3) The cost of the application.

The practicality of each potential application must be evaluated on these three criteria before the method can be implemented.

To provide potential users with an understanding of this evaluation process, factors directly affecting sample size requirements of the genetic method are discussed in this report. In addition, data gathered throughout FY80 were used to study the cost of applying the method and the time required to provide estimates.

## METHODS

### Concept of the Genetic Identification Method

The genetic identification method is based on electrophoretically detected genetic variation (Milner and Tee1 1979; Milner et al. 1980). First, estimates of genotype frequencies are obtained for major stocks that could contribute to a specific fishery; these data are referred to as baseline data. Genotype frequencies are also obtained from a sample of the mixed fishery. The two sets of genotype frequencies are then used to obtain maximum likelihood estimates of the proportional contributions of the stocks to the mixed fishery.

## Electrophoresis

Genetically controlled protein variation was detected by starch gel electrophoresis coupled with histochemical staining [for details on the procedures see May (1975) and Milner et al. (1980)J. Electrophoresis was performed on liver, heart, skeletal muscle, and eye extracts from whole fish samples.

## Maximum Likelihood Analysis

The maximum likelihood estimates of percent contributions were obtained using the EM algorithm (Dempster et al. 1977),

$$\theta^*_j = \sum_{i=1}^{g} \frac{Y_i \, X_{ij} \, \theta_j}{\left(\sum_i Y_i\right)\left(\sum_j X_{ij} \, \theta_j\right)}$$

where yi (i = 1, 2,..., g) = number of fish in the mixed fishery sample having the ith genotype.

$\theta_j$ (j = 1, 2,..., n) = proportion of the mixed fishery from the jth stock.

$x_{ij}$ = frequency of the ith genotype in the jth stock contributing to the mixed fishery.

The estimates are obtained by the following stepwise procedure:

1) $\theta^*_j$ values are obtained through initially solving the equation by assigning arbitrary values to $\theta_j$.

2) A new set of $\theta^*_j$ is then obtained by substituting the previously obtained $\theta_j$ values for $\theta_j$, and again solving the equation.

3) Steps 1 and 2 are repeated until the successive values of $\theta^*_j$ converge.

The formula used to calculate the variance (V) of the ML estimate was:

$$\frac{1}{V_{\theta_{jk}}} = \sum_{i=1}^{g} \frac{Y_i \; (X_{ij} - X_{in}) \; (X_{ik} - X_{in})}{\left[ \sum_{j}^{n-1} X_{ij}\theta_j + X_{in} \; (1 - \sum_{j}^{n-1} \theta_j) \right]^2}; \; k = 1, 2, \ldots, n \text{ stocks.}$$

where all terms of this expression are defined as above.[1/]

Baseline genotype frequencies ($Xi_j$) were considered fixed constants. Therefore, variances calculated with the formula given above do not include variation associated with the estimates of baseline genotype frequencies. To include this source of variation in the calculations would be too difficult and costly to be practical. Rather, it is better to minimize this source of variation by using adequate sample sizes for estimating the baseline genotype frequencies.

## Baseline Data

USFWS collected and immediately froze 200 spring chinook from each of the following Columbia River hatcheries: Rapid River (RR), Kooskia (KO), Leavenworth (LW), Round Butte (RB), Warm Springs (WS), Klickitat (KT), Little White Salmon (LS), Carson (CA), Oakridge (OR), McKenzie (MK), South Santiam (SS), Eagle Creek (EC), Kalama (KA), and Cowlitz (CO) (Figure 1). These samples were shipped to the genetics laboratory at the NMFS' Manchester Marine Experimental Station, Manchester, Washington, for electrophoretic analysis. Data from 10 loci were obtained from these samples and were used as baseline data in the analyses of the blind tests and simulated mixed fisheries (Appendix A).

1/ This is a correction of the formula given in Milner and Tee1 (1979) and in Milner et al. (1980).

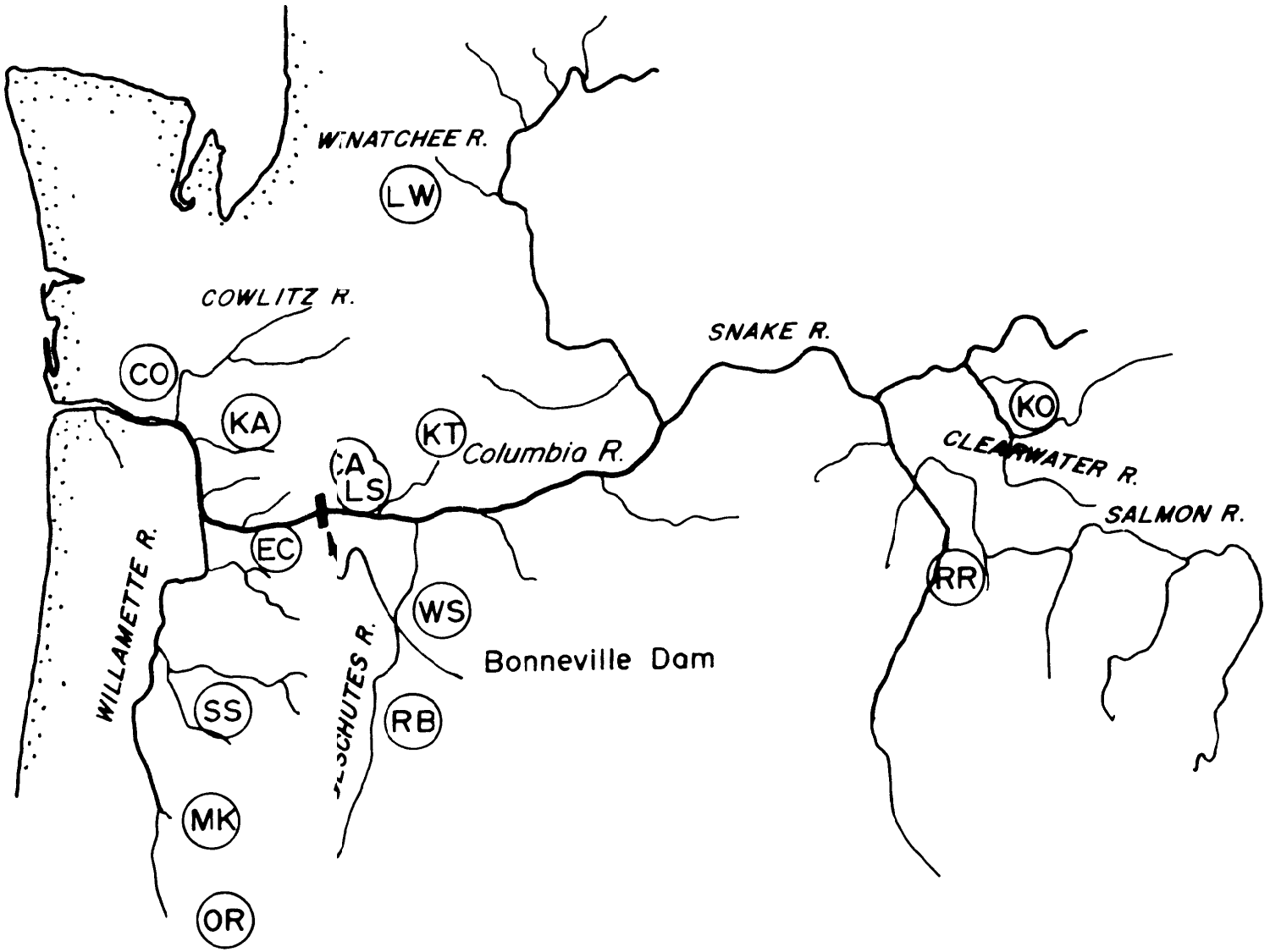


Figure 1.--Spring chinook salmon stocks used in blind test demonstration (abbreviations are defined in the text).

### Blind Samples A, B, C, and D

As mentioned previously, USFWS also collected and froze additional fish from the 14 hatcheries. From these, four samples with different compositions were constructed (see Appendix B). The actual proportions of the contributing stocks were not known by NMFS personnel. The four mixed samples were shipped to Manchester and electrophoretically assayed for the same 10 polymorphic loci as the baseline samples. Estimates were made of the percent contribution of each stock and each geographic group of stocks to the four blind samples. USFWS then revealed their actual compositions for comparison with the estimates.

### Relationship of Calculated to Observed Standard Deviation

A study was conducted to determine how standard deviations obtained from the formula for calculating variances compared with actual standard deviations. The composition of the fishery obtained by combining all four blind samples was estimated from 10 or more random samples of 100, 200, 500, and 1,000 individual-s. Ratios of the calculated standard deviations to the observed standard deviations were calculated for 14 stocks and 5 groups of stocks for each sample size. The relationship of these ratios to sample size was examined.

### Accuracy and Precision: An Effect of Potential Contributors

Two fisheries (A and B) were simulated to demonstrate that the accuracy and precision with which a stock is estimated depends upon which other stocks are believed to be potential contributors. The actual composition of both fisheries was defined to consist of only two stocks

(Kooskia and Rapid River). The percent contribution of each of these stocks was approximately equal in both fisheries.

It was assumed that 14 stocks could potentially contribute to Fishery A but only two stocks to Fishery B. Therefore, the compositions of the two fisheries were estimated using the 14 stock baseline (including KO and RR) for Fishery A and a two stock baseline (KO and RR) for Fishery B.

## Lower Columbia Gill-Net Fishery: A Field Test

On 27 and 28 February 1980, a brief commercial fishery on the spring chinook salmon run of the lower Columbia River provided an opportunity to field test the method. The field test was necessary to collect time and cost data as well as to gain valuable experience in applying the method to an actual fishery.

Personnel from Washington State Department of Fisheries (WDF) and Oregon Department of Fish and Wildlife (ODFW) provided a list of licensed buyers which historically processed the greatest number of fish and also were cooperative with previous sampling efforts. The following buyers were contacted and granted permission to the NMFS sampling crew to take tissue samples from their chinook salmon:

| | |
|---|---|
| Astoria Fish Factors Inc. | Astoria, Oregon |
| Astoria Seafood | Astoria, Oregon |
| Barbey Packing Corp. | Astoria, Oregon |
| Bumble Bee Seafoods | Astoria, Oregon |
| Josephson's Smokehouse & Dock | Astoria, Oregon |
| Chinook Packing Co. | Ilwaco, Washington |
| Jessie's Ilwaco Fish Co. | Ilwaco, Washington |
| Whitney Fidalgo Seafoods Inc. | Seattle, Washington |

ODFW and WDF estimated a total catch of 305 chinook salmon in the 24-hour fishery. The NMFS crew sampled 123 of the fish. Heart and liver tissues were taken as the fish were dressed at the dressing stations. Sampling of eye and skeletal muscle tissues was not permitted by the buyers.

The 123 samples were immediately frozen and shipped to Manchester for electrophoretic analyses. The use of heart and liver tissues allowed the collection of data for nine polymorphic loci. All Columbia and Snake River spring chinook stocks were considered to be potential contributors to this fishery. Therefore, the baseline used included the complete set of stocks (Table 1) for which data for the nine polymorphic loci were available. Estimates of the composition of the fishery were made and the interested management agencies informed.

## Time and Cost Study

To provide guidelines for estimating the turnaround time and cost of potential applications, a time and cost study was conducted throughout FY80. The data gathered were used to create an exemplary time and cost schedule for a hypothetical application of the method.

This example was based on the following conditions:

(1) The specific management question to be addressed required a sample from the fishery of 500 individuals.

(2) The sampling was done in dressing stations within 5 hours shipping time of the genetics laboratory.

(3) Sufficient numbers of fish were dressed to allow maximum productivity by the sampling crews.

Table 1.--Baseline Stocks for Winter Fishery.

SNAKE RIVER:

- Rapid River
- Kooskia
- Grande Ronde
- Imnaha
- Mid-Fork of the Clearwater

UPPER COLUMBIA RIVER (Above confluence of Snake River):

- Leavenworth
- Wenatchee
- Twisp
- Entiat
- Yakima

BONNEVILLE DAM TO MCNARY DAM:

- John Day
- Round Butte
- Warm Springs
- Klickitat
- Carson

BELOW BONNEVILLE (Non-Willamette River):

- Lewis
- Kalama
- Cowlitz

WILLAMETTE RIVER:

- Eagle Creek
- South Santiam
- North Santiam
- Dexter
- Middle Fork Willamette
- McKenzie

(4) Trained personnel included:

a. A sampling crew of six technicians.

b. A laboratory staff of seven technicians.

c. A computer technician.

d. A project leader.

(5) Electrophoretic assays were for the 10 polymorphic loci currently used in mixed fishery analyses of chinook salmon.

(6) Costs to NMFS were based on the present economy.

The resulting time and cost schedule can provide guidelines for determining the practicality of a potential application of the method.

## RESULTS

The following results, and in particular those of the blind sample study, confirm that the genetic identification method is an effective means for identifying components of mixed stocks. However as with any system of estimation, the estimates are meaningful only when their variation is considered. Therefore, the estimates presented in this report are considered, where appropriate, from two viewpoints: accuracy and precision.

Accuracy describes the closeness of an estimate to its true value. Therefore, the difference between the actual and estimated percent contribution (D) is used to evaluate the accuracy of an estimate.

Precision is concerned with the repeatability of an estimate and is closely related to variance and standard deviation. In this report, the standard deviation (SD) of an estimate is used to measure precision.

## Blind Samples A, B, C, and D

The results obtained for blind samples A, B, C, and D of the controlled test demonstration are presented at three levels of detail: (1) 14 stocks, (2) five geographic groups, and (3) two geographic groups (See Table 2). These results are shown graphically in Figures 2a-d and are tabulated in Appendix B. The accuracy of an estimate is indicated by the distance from the circle to the horizontal bar and precision by the length of the vertical line.

These graphs and tabulated data show that the precision and accuracy of the estimates varied considerably among the stocks and geographic groups. Yet it can also be seen that the relative magnitudes of the SD's between stocks and between groups were consistent over the four samples, e.g., the estimates for Rapid River were invariably more precise than those for Kooskia. Variation in accuracy fits no obvious pattern (as expected for single point estimates). However, it is noteworthy that 64% of the estimates were zero and 80% were within 2.3 percentage points when the actual contribution was zero.

The data presented in Table 3 allows a more quantitative evaluation of the results of the blind sample study. The average difference in percentage between the estimate and actual value (D) over the four blind samples was 3.4, 3.2, and 2.3% for 14 stocks and five and two geographical groups, respectively. The average D values ranged from 0.9 to 7.4%, while individual D values varied from 0.0 to 18.4%. Larger D's tended to be associated with the larger estimated contributions. The mean standard deviation over the four samples averaged 4.3, 5.6, and 1.0% and ranged from

Table 2.--Stocks and geographic groupings of blind samples.

| Abbreviation | Stock | Geographic Group | |
|---|---|---|---|
| RR | Rapid River | SN - Snake River | |
| KO | Kooskia | | |
| LW | Leavenworth | UC - Upper Columbia (above the confluence of the Snake River) | |
| RB | Round Butte | BP - Bonneville Dam to McNary Dam | AB - Above Bonneville |
| WS | Warm Springs | | |
| KT | Klickitat | | |
| LS | Little White Salmon | | |
| CA | Carson | | |
| OR | Oakridge | W - Willamette | BB - Below Bonneville |
| MK | McKenzie | | |
| SS | South Santiam | | |
| EC | Eagle Creek | | |
| KA | Kalama | K/C - Kalama/Cowlitz | |
| CO | Cowlitz | | |

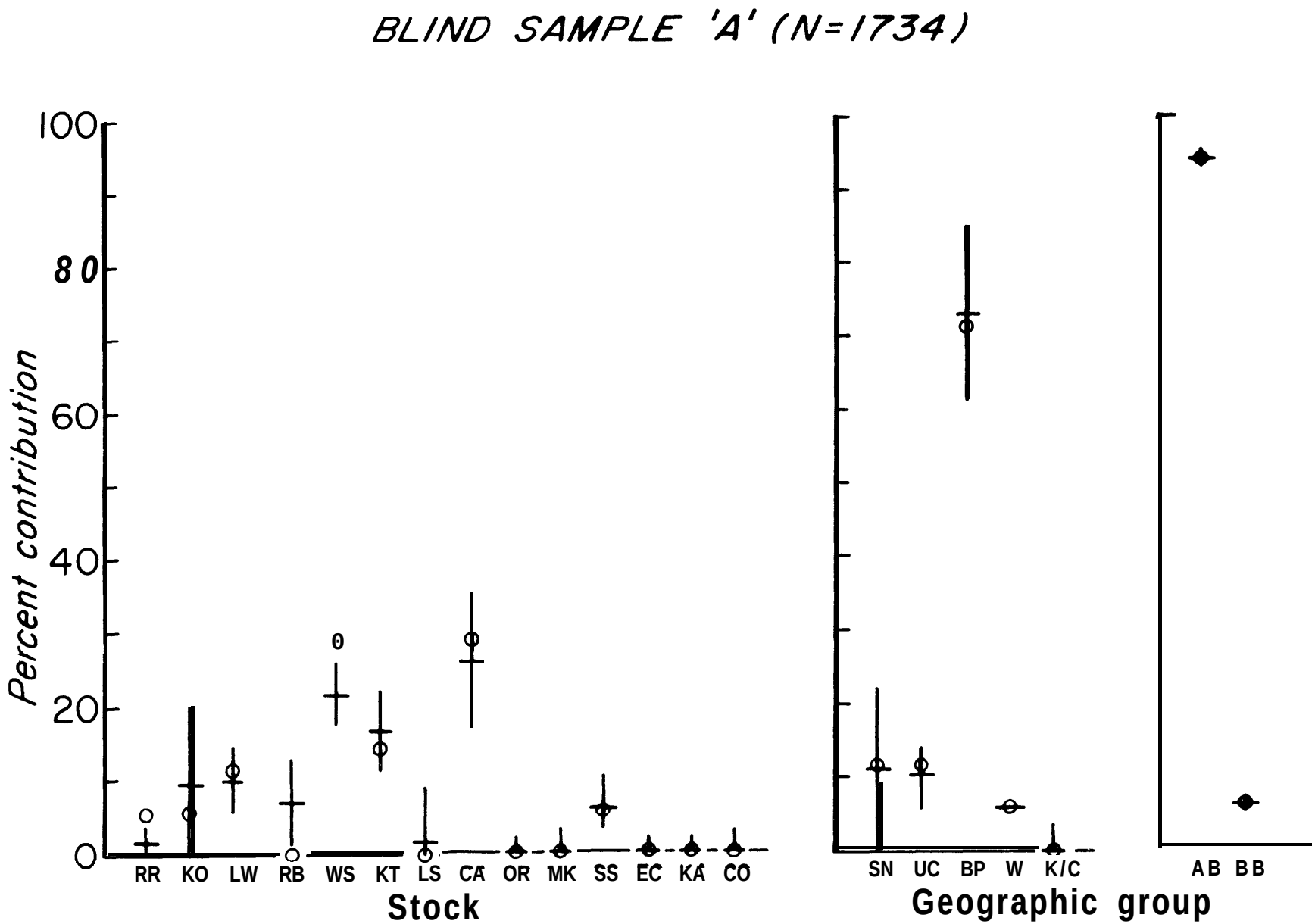


Figure 2a.--Estimated contributions for blind samples (A) of spring chinook salmon. Circle = actual contribution, horizontal bar = estimated contribution, and vertical line = ± one standard deviation.

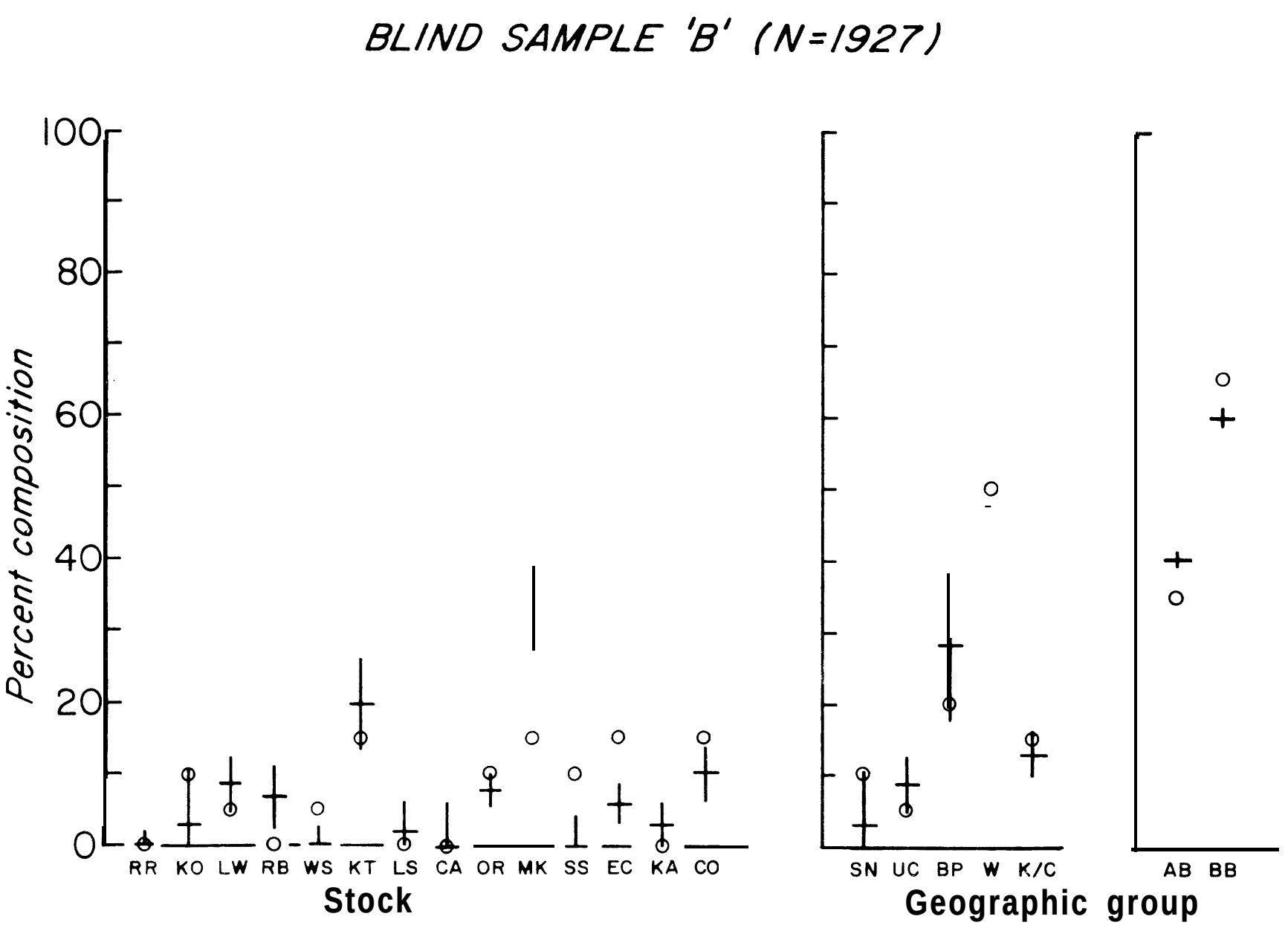


Figure 2b. --Estimated contributions for blind samples (B) of spring chinook salmon. Circle = actual contribution, horizontal bar = estimated contribution, and vertical line = ± one standard deviation.

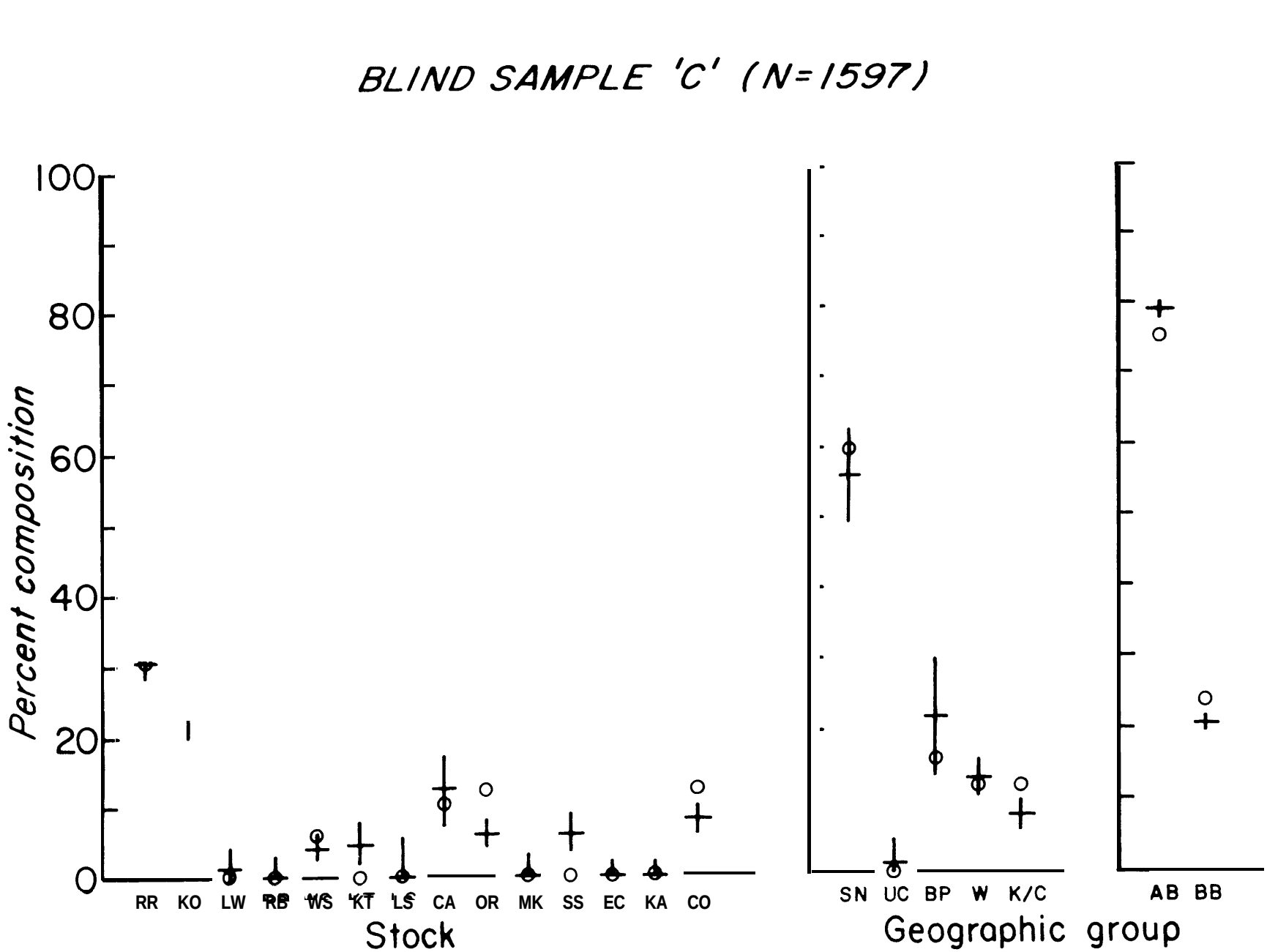


Figure 2c.--Estimated contributions for blind samples (C) of spring chinook salmon. Circle = actual contribution, horizontal bar = estimated contribution, and vertical line = ± one standard deviation.

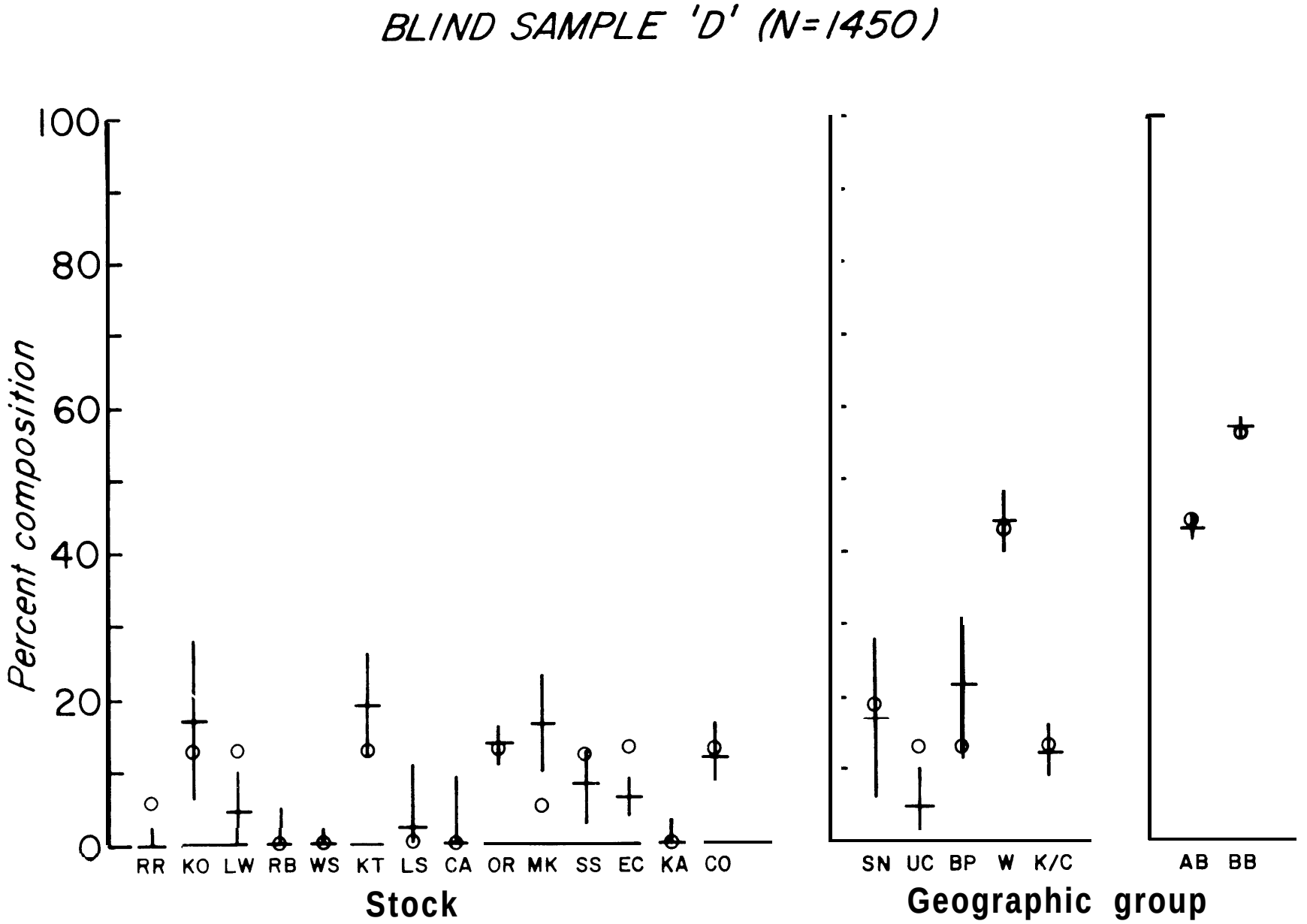


Figure 2d. --Estimated contributions for blind samples (D) of spring chinook salmon. Circle = actual contribution, horizontal bar = estimated contribution, and vertical line = ± one standard deviation.

Table 3.--Estimated contribution, standard deviation of the estimate, and difference between actual and estimated contributions averaged over four mixed stock samples having known compositions.

| Contributor | Estimated contribution | Standard deviation of the estimate | Difference between actual and estimated contribution [1/] |
|---|---|---|---|
| Stock (14) | | | |
| Rapid River | 7.9 | 2.2 | 2.6 |
| Kooskia | 13.9 | 9.1 | 4.7 |
| Leavenworth | 6.1 | 4.1 | 3.7 |
| Round Butte | 3.4 | 4.4 | 3.4 |
| Warm Springs | 6.6 | 2.6 | 3.3 |
| Klickitat | 15.2 | 5.4 | 4.6 |
| Little White Salmon | 1.3 | 6.2 | 1.3 |
| Carson | 9.7 | 7.7 | 1.2 |
| Oakridge | 7.1 | 2.4 | 2.4 |
| McKenzie | 12.5 | 4.9 | 7.4 |
| Youth Santiam | 4.9 | 4.0 | 5.4 |
| Eagle Creek | 3.2 | 2.1 | 3.8 |
| Kalama | 0.8 | 2.5 | 0.9 |
| Cowlitz | 7.5 | 3.1 | 2.5 |
| Average | 7.7 | 4.3 | 3.4 |
| Geographic group (5) | | | |
| Snake | 21.7 | 9.3 | 3.2 |
| Upper Columbia | 6.1 | 4.1 | 3.7 |
| Bonneville Dam to McNary Dam | 36.1 | 10.0 | 6.0 |
| Willamette | 27.6 | 1.7 | 1.4 |
| Kalama/Cowlitz | 28.3 | 3.0 | 1.7 |
| Average | 20.0 | 5.6 | 3.2 |
| Geographic group (2) | | | |
| Above Bonneville | 63.9 | 1.0 | 2.3 |
| Below Bonneville | 36.0 | 1.0 | 2.3 |
| Average | 49.9 | 1.0 | 2.3 |

1/ Only the magnitudes of the differences were used to caculate these averages.

2.1-9.1, 1.7-10.0, and 1.0-l 3% for 14 stocks and five and two geographic groups.

A definite gain in precision relative to estimated contribution size and a slight gain in accuracy were achieved on the average when the 14 stocks were pooled into geographic groups. Specifically, average reductions in the ratio of SD to estimated percent contribution (S/E) of 49.8% and in D of 5.9% were obtained when the 14 stocks were pooled into five groups. When the five groups were pooled into two groups even greater average reductions in S/E and in D were obtained. The reduction in S/E was 92.3% and in D 28.1%.

## Relationship of Calculated to Actual Standard Deviation

The ratios of calculated standard deviations to actual standard deviations ($SD_C/SD_A$) were calculated from the results of this study. Figures 3a-d present the ratios for 14 stocks and five groups of stocks plotted against the mean estimated contribution for the four sample sizes. The ratio was approximately independent of combined effects of contribution and stock or group of stocks. One exception was at contributions of less than 1% where the ratio was always excessively higher than the average. This indicated that the standard deviation that was calculated by formula under this condition fails to provide a reasonable approximation of the true variance. Figure 4 is a plot of the ratio $SD_C/SD_A$ averaged over individual and geographic groups of stocks as a function of sample size. The average ratio ranged from 3.0 for N = 100 to 1.3 for N = 1000 and displayed a simple linear relationship ($r = -0.99$) to sample size when plotted logarithmically. Based on this relationship, the two SD's became equal (i.e., $S_C/S_A = 1.0$) when N is extrapolated to 1940.

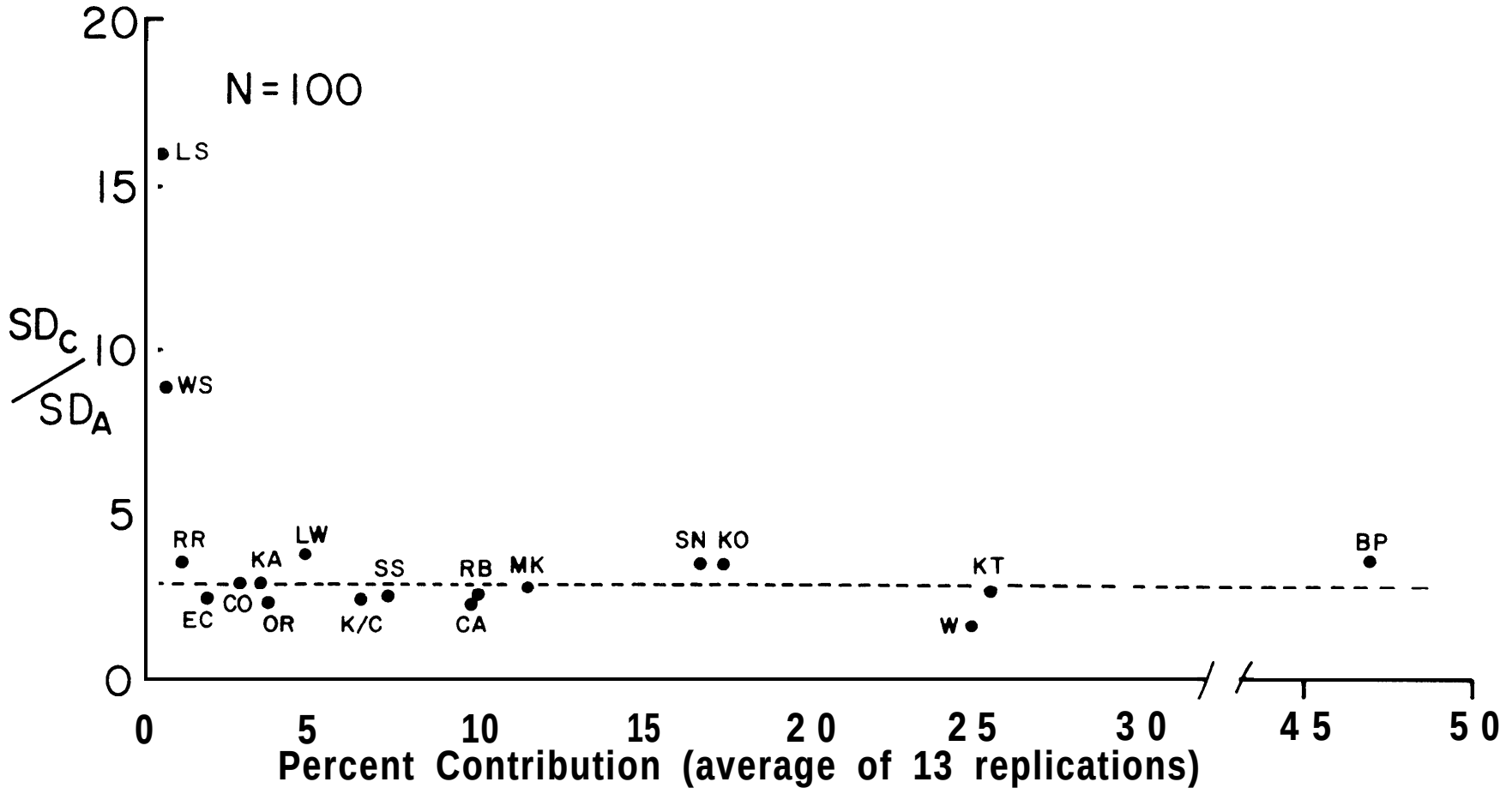


Figure 3a.--The ratios of formula calculated to actual (observed) standard deviations plotted against estimated contributions for sample size 100. The broken horizontal line indicates the mean value (excluding contributions of less than one percent) of the ratio.

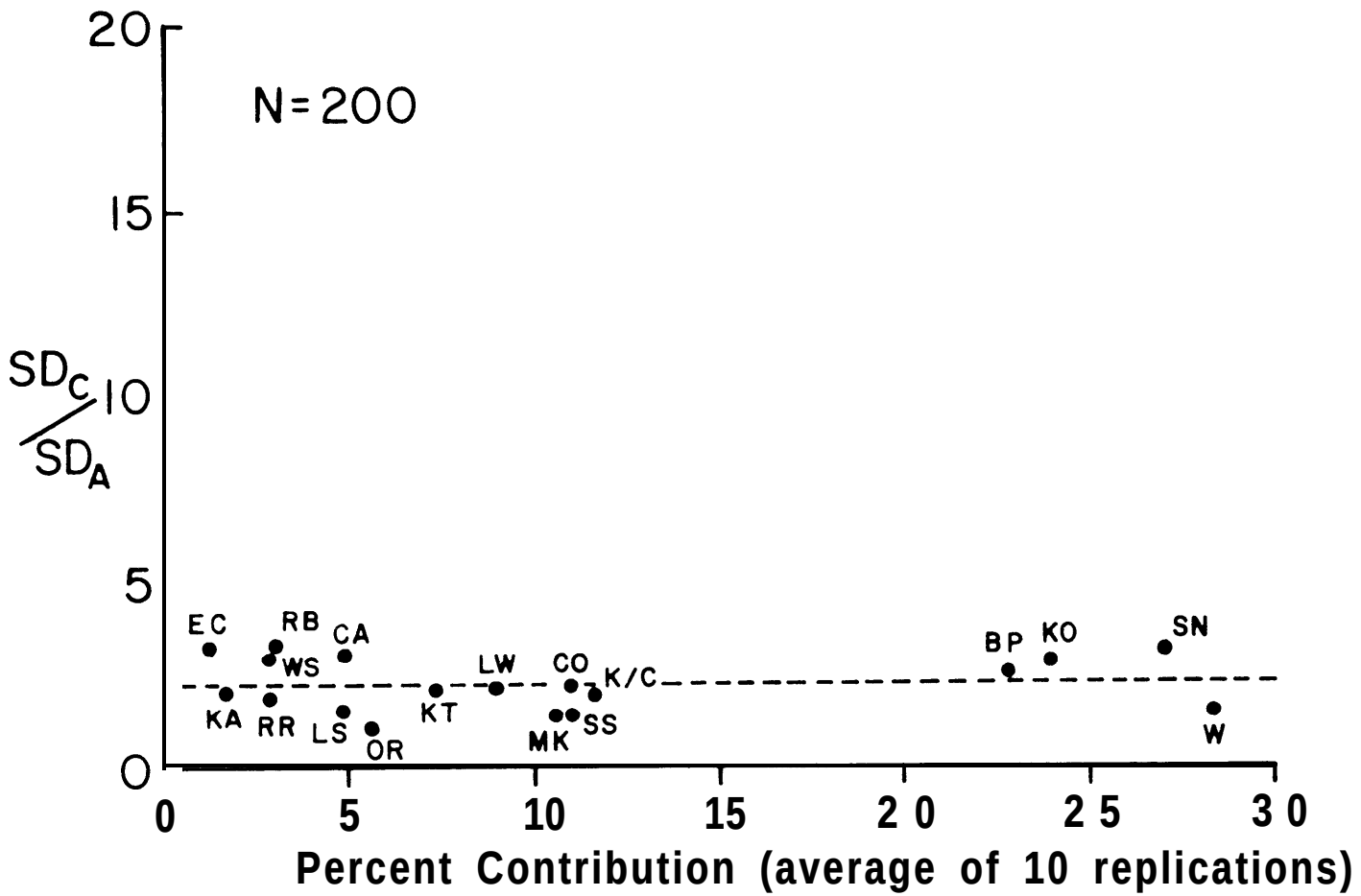


Figure 3b.--The ratios of formula calculated to actual (observed) standard deviations plotted against estimated contributions for sample size 200. The broken horizontal line indicates the mean value (excluding contributions of less than one percent) of the ratio.

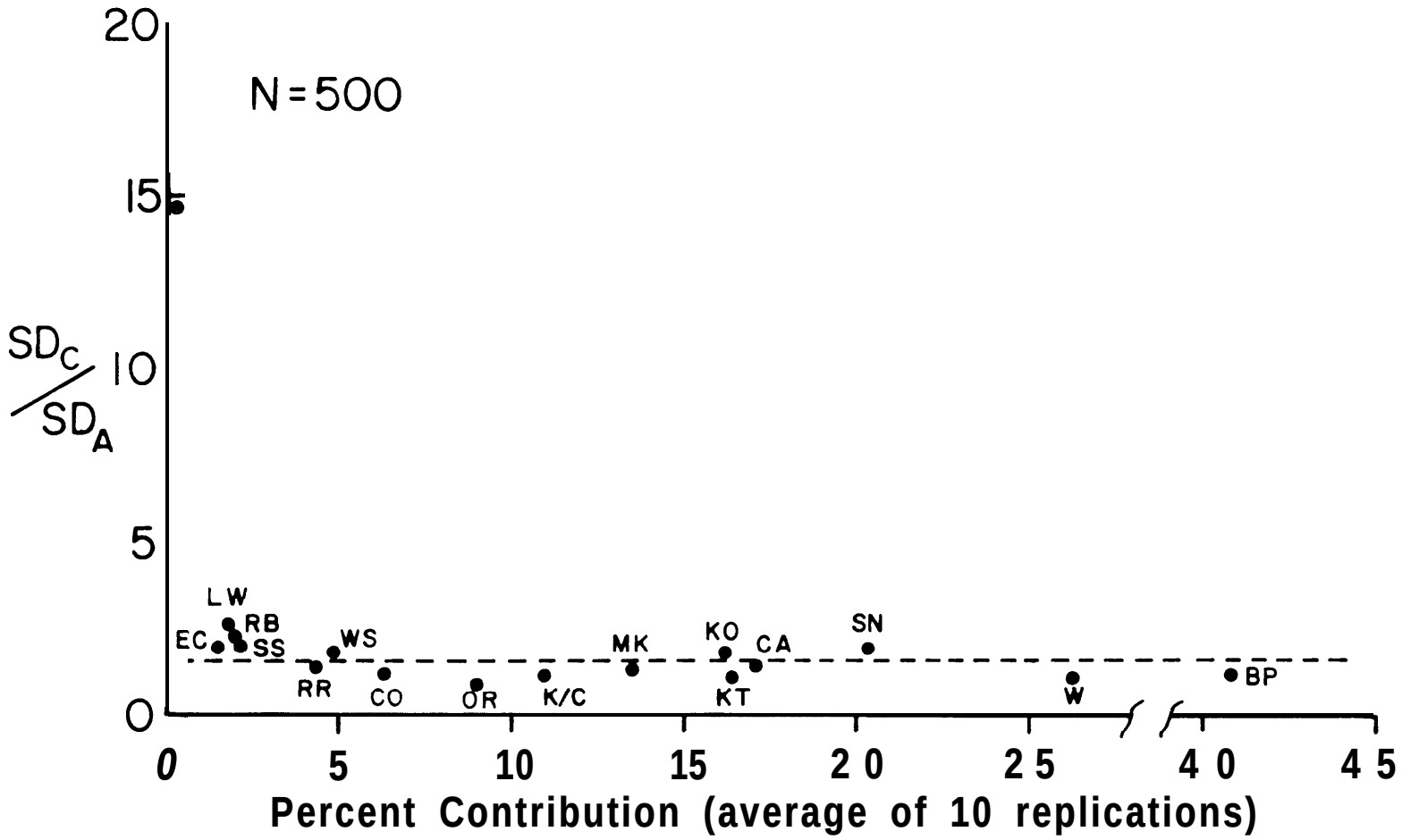


Figure 3c. --The ratios of formula calculated to actual (observed) standard deviations plotted against estimated contributions for sample size 500. The broken horizontal line indicates the mean value (excluding contributions of less than one percent) of the ratio.

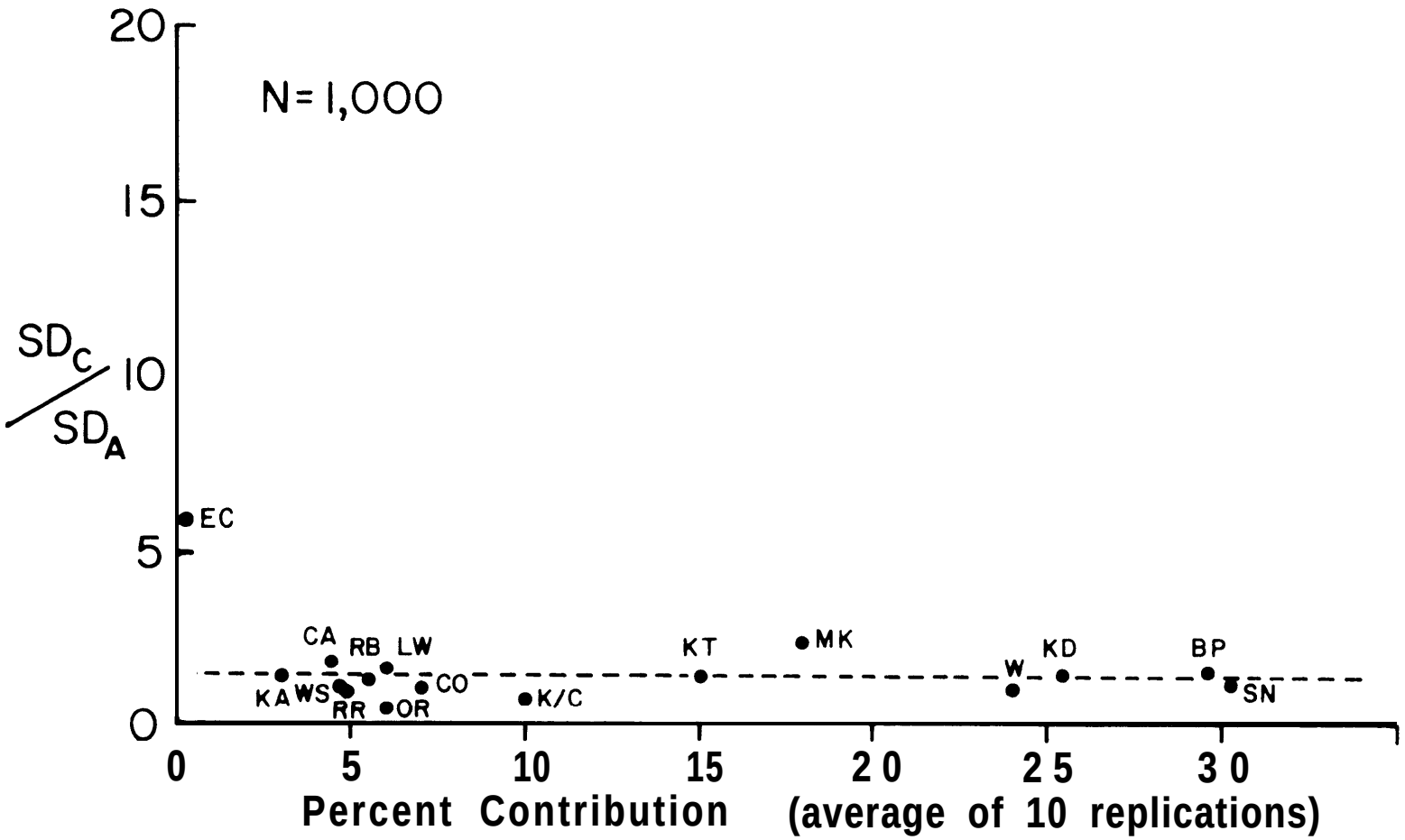


Figure 3d.--The ratios of formula calculated to actual (observed) standard deviations plotted against estimated contributions for sample size 1,000. The broken horizontal line indicates the mean value (excluding contributions of less than one percent) of the ratio.

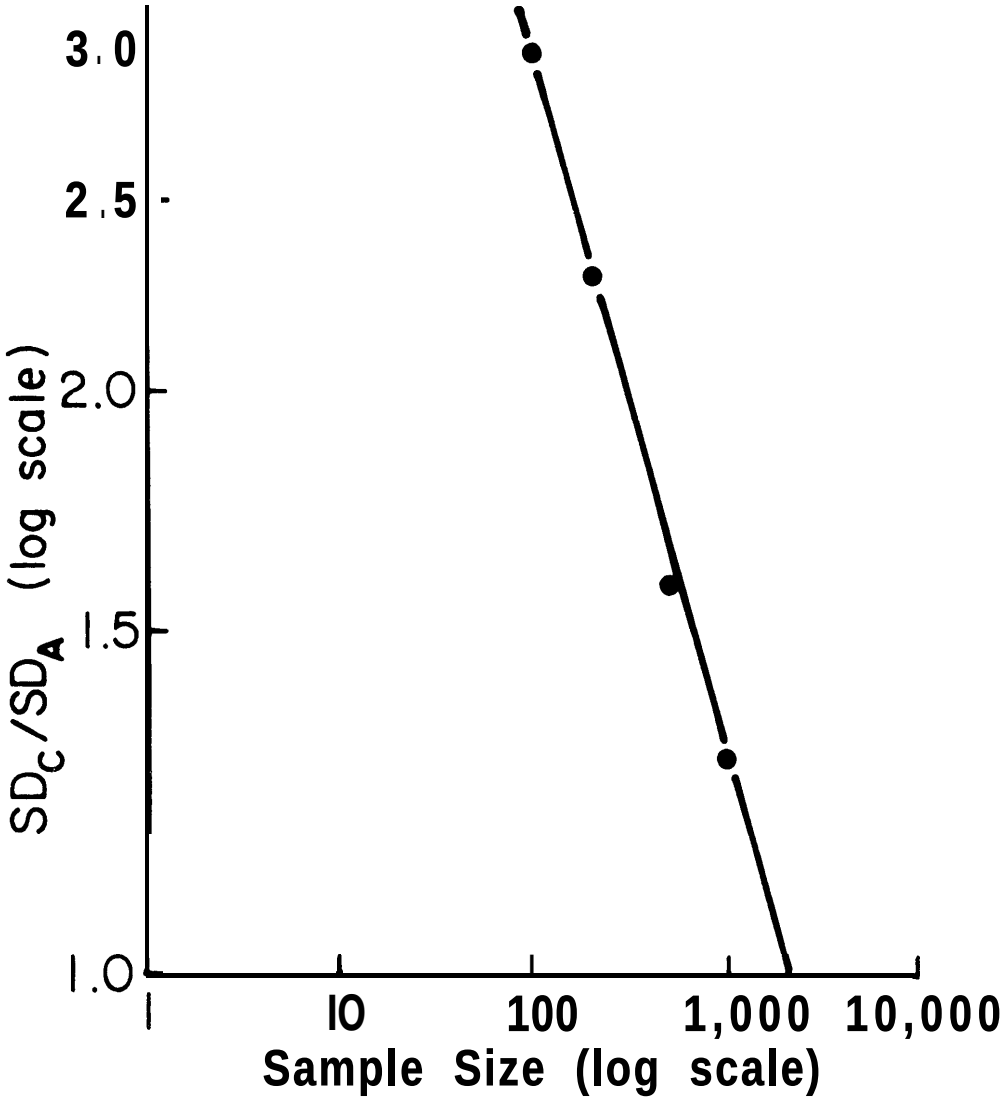


Figure 4.--Loglinear regression of the ratio of formula calculated to actual (observed) standard deviations on sample size.

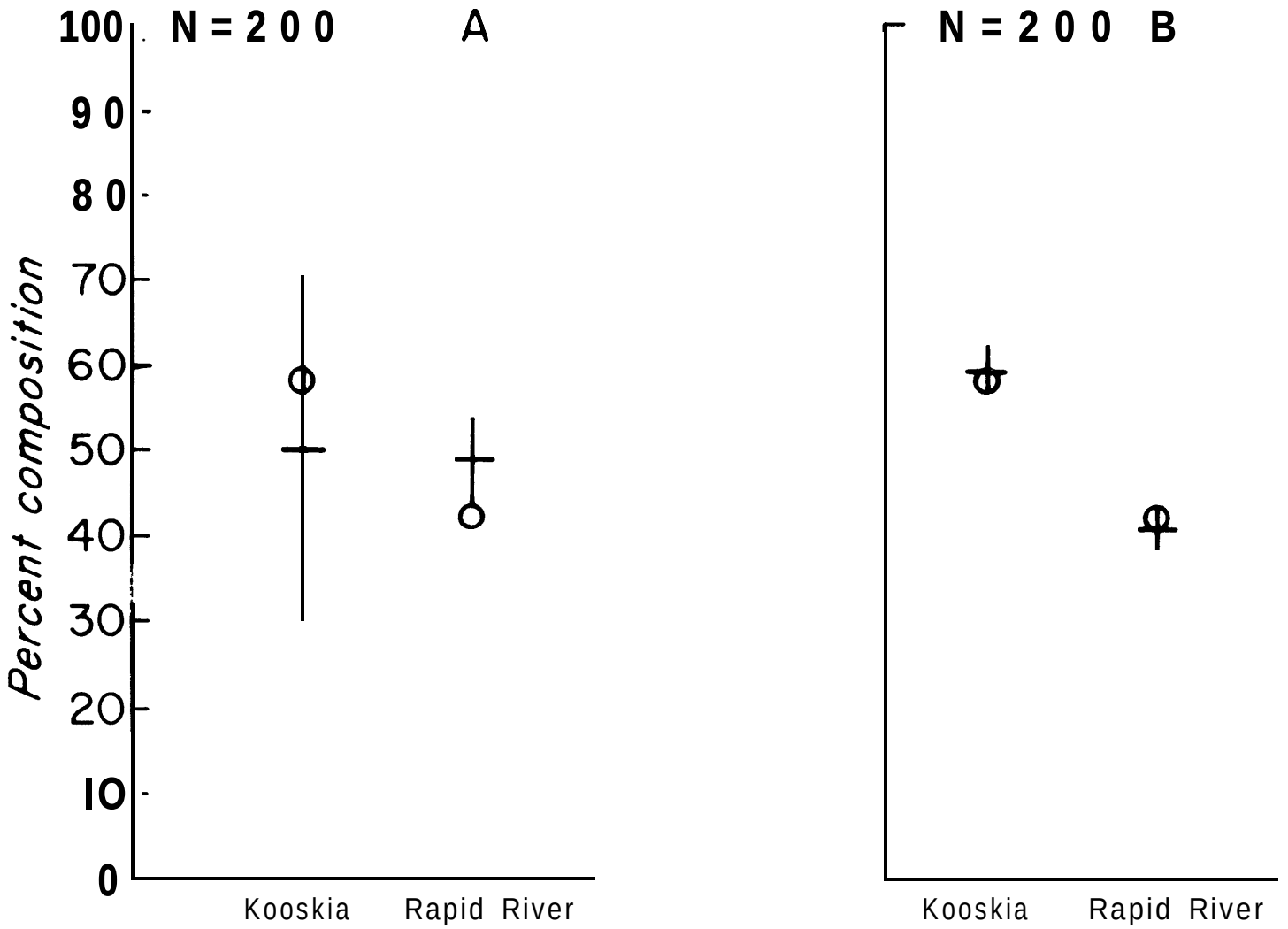


Figure 5.--Estimated contributions of two spring chinook salmon stocks in two fisheries. A (left) 14 potential contributors and B (right) 2 potential contributors. Circle = actual contribution, horizontal bar = estimated contribution, and vertical line = ± one standard deviation.

### Accuracy and Precision: An Effect of Potential Contributors

The results of this study clearly demonstrate that the accuracy and precision with which a stock is estimated depends upon which other stocks possibly contribute to the fisheries (Figure 5). D values for Fishery A (14 potential contributors) were 7.5% for Kooskia and 6.5% for Rapid River. These values were reduced to 1.1% for both Kooskia and Rapid Rivers for Fishery B (2 potential contributors). The SD for Kooskia for Fishery A was 19.6% and only 3.3% for Fishery B. The SD for Rapid River for Fishery A was 4.8% and 3.3% for Fishery B.

### Lower Columbia Spring Chinook Gill-Net Fishery

The estimated stock composition of the spring chinook salmon harvested in the winter gill-net fishery was 87% from below Bonneville (BB) and 13% from above Bonneville (AB). Willamette stocks accounted for an estimated 100% of the BB contribution to the harvest. Columbia and Snake River stocks were estimated to have made approximately equal contributions to the AB harvest.

The reliability or precision of the estimates was low due to the small sample size and to the assumed complexity of the fishery (24 potentially contributing stocks); estimated zero contributions were obtained for 18 of the 24 stocks. This indicates that perhaps the fishery was assumed to be more complex than necessary. If the complexity could have been reduced, preferably on an a priori basis, the precision of the estimates would have been increased. Nevertheless, a meaningful 68% confidence interval of 72-100 was obtained for the estimated percent contribution of the Willamette stocks as a group.

## Time and Cost Estimates for Hypothetical Application

Data gathered throughout 1980 was used to create a time and cost schedule for a hypothetical application of the method. The conditions of the hypothetical application are described in the Methods section of this report. Figure 6 presents the time schedule, and Table 4 presents results of the cost study.

Cost results were based on the collection and analysis of 500 samples within a 24-hour period. The requirement of an intensive effort by an experienced staff (one project leader, seven lab technicians, six sample collectors, and one computer technician) resulted in salaries and overhead costs totaling $4,156 or $8.31 per fish. The transportation of sampling personnel to and from the sampling site and samples to the lab, cost $525 or $1.05 per fish. Miscellaneous supplies cost $465 or $0.93 per fish. The total cost per fish was $10.29.

Time results were for the collection, transportation, lab processing, and analysis of 500 samples. Sample collection is the bagging, labeling, and freezing of portions of tissue as the fish are being dressed. An estimated 3 hours were required by a crew of six technicians to collect the samples. The transportation of the frozen samples from the sampling site to the lab required 5 hours. Sample preparation is the subsampling of the tissues and extraction of the proteins. An estimated 4 hours were required for sample preparation. Laboratory preparation includes the readying of all materials and equipment and was estimated to take 3 hours. Electrophoresis, the technique of assaying for genetically controlled protein variation, required an estimated 9.5 hours. Data recording is the numerical coding of the genetic data and was estimated to take 6.5

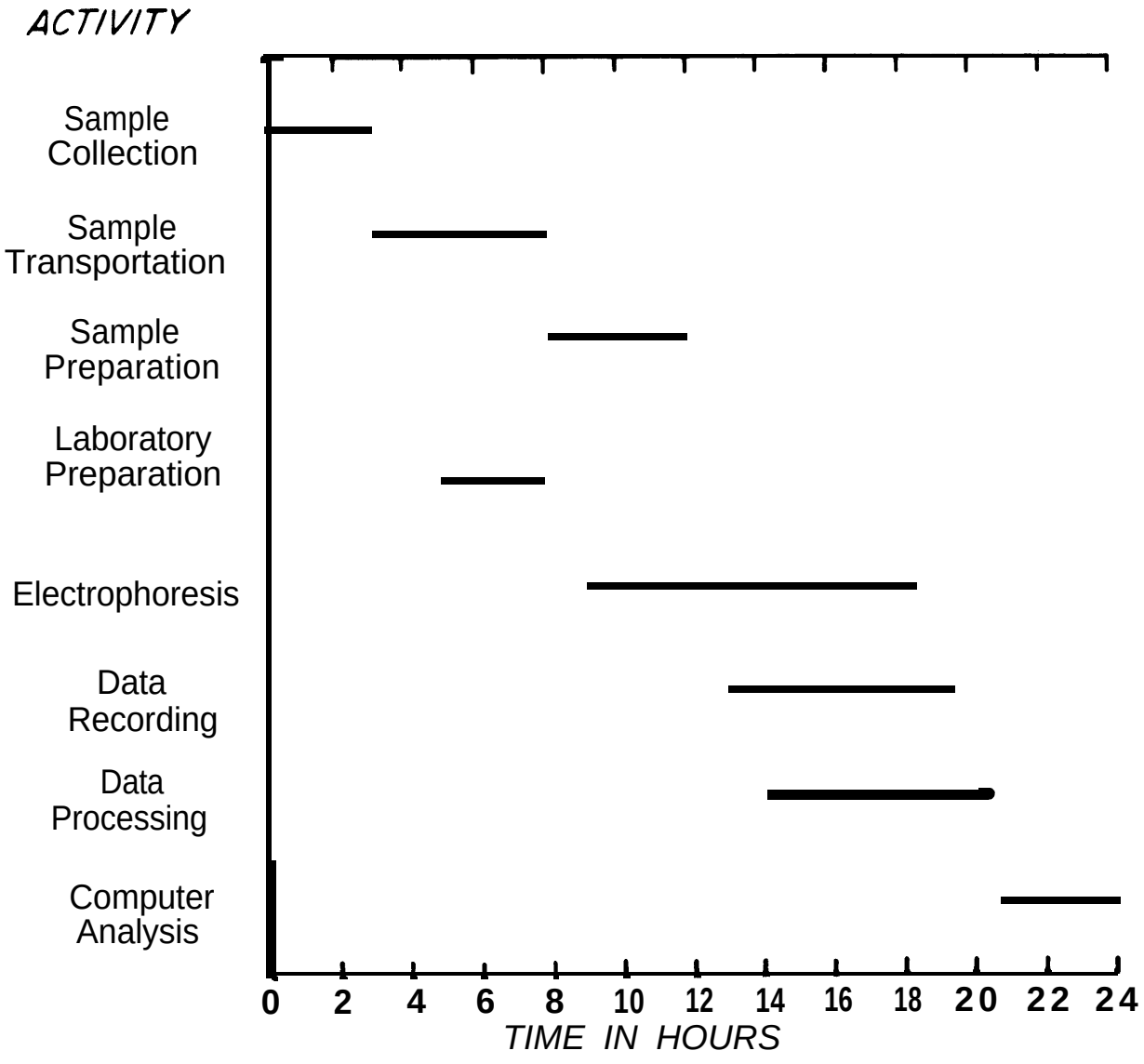


Figure 6.--Exemplary time schedule for a mixed fishery analysis. (Sample size = 500)

Table 4.--Cost Estimate for a Mixed Fishery Analysis.

| | | | |
|---|---|---|---|
| **Salaries** | | | |
| 1 Project Leader | GS-11 | 605 | |
| 4 Lab Technicians | GS- 5 | 330 | |
| 3 Lab Technicians | GS- 7 | 350 | |
| 6 Sample Collectors | GS- 5 | 566 | |
| 1 Computer Technician | GS- 7 | 58 | |
| Total Base Pay | | 1,909 | |
| Leave Surcharge (16.9% of Base Pay) | | 323 | |
| Subtotal | | 2,232 | |
| Overtime | | 415 | |
| Total Direct Labor | | 2,647 | |
| Employer's Contribution (9.6% of Subtotal) | | 214 | |
| Total Salaries | | | 2,861 |
| **Operations and Supplies** | | | |
| Travel | | 525 | |
| Miscellaneous Supplies | | 465 | |
| Total Operations and Supplies | | | 990 |
| **Support Costs** | | | |
| S.L.U.C. | | 139 | |
| NOAA Support (42.7% Total Direct Labor) | | 1,130 | |
| DOC (0.5% of Total Costs) | | 26 | |
| Total Support | | | 1,295 |
| Estimated Cost | | | $5,146 |

hours. Data processing is the entering of the data into the computer system and required an estimated 6.5 hours. Computer analysis includes data file management, program execution, and interpretation of the resulting estimates. An estimated 3.5 hours were required for computer analysis. The efficient scheduling of these activities resulted in a total turnaround time of 24 hours.

## DISCUSSION

The reliability of the genetic identification method can be evaluated by examining the relationship between the precision or repeatability of the estimates and the amount of genetic difference between stocks. Because the method uses genetic differences between stocks to estimate their proportions in mixtures, stocks that are more different than others should be estimated more precisely.

There is, in fact, considerable variation in the amount of genetic difference between the 14 stocks used In the blind sample study. These genetic differences [we use Nei's (1974) measure] have occurred naturally and as expected often parallel the geographic structure of the river system. For example, the Rapid River stock is of Snake River origin and very different from the other 13 stocks. The average genetic similarity between Rapid River and the other stocks is 0.975. In contrast, the Oakridge, McKenzie, South Santiam, and Eagle Creek stocks of the Willamette River have a higher average genetic similarity of 0.996.

However, some deviations from this parallel genetic and geographic structuring has resulted from the transplantation of stocks. For example, the Kooskia stock of the Snake River has been extensively supplemented with stocks of Columbia River origin; principally Carson (which in turn was

derived from a mixture of Snake River and Columbia River fish). Consequently, the genetic similarity between the Kooskia and Carson stocks is very high (0.9997).

If the method of estimation is working correctly, it is expected that the relative precision of the estimates would be consistent with the degree of genetic difference. Estimates for Rapid River would be more precise (have smaller SDs) than those for Kooskia and Carson. This expectation is, in fact, consistently the case as can be seen in Figures 2a-d and in Table 2.

It is also expected that the relative precision of pooled estimates would be consistent with the pattern of genetic differences. Geographic groupings, which combine estimates with high genetic similarities would result In reduced SDs. The precision of the estimates for the Willamette stocks is evidence that this is true. SDs for the individual estimates for these four stocks average 3.3% over the four blind samples. When the estimates are pooled, the average SD for the Willamette group is only 1.7%.

It is noteworthy that the average SDs for the Snake (9.3%) and Bonneville Dam to McNary Dam (10.0%) groups are much larger than that for the Willamette group. This difference is in part due to the high genetic similarity between Kooskia of the Snake group and Carson of the Bonneville pool group. Only when the geographic grouping is carried a step further, and the estimates for all stocks above Bonneville (including Kooskia and Carson) are combined, does the SD become very small (1.0%). The consistent pattern of the precision of the estimates in the blind sample study verify the reliability of the method.

The practicality of a potential application of the method is largely determined by the number of fish to be sampled from the fishery. Sample

size requirements not only affect turnaround time and cost, but also must be reasonable relative to the value of the information that is obtained. Estimating the required sample size, turnaround time, and cost is therefore the initial step in using the method.

Planning begins with the identification of the stock(s) or group(s) of stocks in the fishery for which estimates are needed. The user of these estimates must also establish the level of precision that is necessary to make the management decision. A computer simulation of the fishery is then used to estimate the sample size that is required.

The size of the sample that is needed is determined by the amount of genetic difference between the stocks to be estimated. Consequently, combined estimates of genetically similar stocks are often more precise than those for individual stocks. This can be an effective strategy for reducing sample size requirements when the estimation of the combined contribution of a group of stocks provides the needed management information. For example, Figures 2a-d show that estimates for the combined above Bonneville stocks were more precise than those for the individual stocks in this group. Thus, a much smaller sample size is required for equally precise estimates.

Another effective means of decreasing sample size requirements is to reduce the complexity of the baseline. Figure 5 shows the dramatic effect of a reduced baseline on the precision of the estimates and, hence, sample size. Where possible, historical or supplemental data can be used to identify stocks which do not contribute significantly to a fishery and eliminate them from the baseline data.

An increase in the amount of genetic difference that can be detected can also substantially reduce sample size requirements (Milner et al.

1980). This increase result. from the addition of discriminating genetic loci to the data base and is an ongoing goal of this project.

The results of the time and cost study provide approximate guidelines for planning applications of the method. The time and cost schedules presented in this report are currently accurate for the present economy and assumed conditions. Adjustments in these schedules are necessary under different conditions.

A major component of the cost per fish is salaries and overhead. These costs are directly affected by the number of fish sampled, the availability of fish, the number of proteins to be assayed, and the turnaround time required. These conditions directly determine the number of trained technicians and the amount of overtime hours that are needed. Because many of the technical activities require skilled workers, the availability of a large workforce for a short period is a possible constraint. However, if a longer time interval is permissible, this constraint is removed and cost per fish can be reduced.

Another component of the cost per fish is the cost of travel and the shipping of samples. In the example, a 5-hour drive from sample site to the lab is assumed. A more remote sampling site may significantly add to these costs.

Similar adjustments in turnaround time are required under varying conditions. Factors which most affect turnaround time are the availability of fish at the sampling site, the time required to ship samples to the lab, the number of proteins to be assayed, and the availability of experienced technicians.

The hypothetical application presented in this report provides an accurate estimate of the cost ($10.29 per fish) and turnaround time (24

hours) of using the method to estimate the composition of a chinook fishery in the lower Columbia River for an arbitrarily chosen sample size of 500 fish.

The data of this report fulfill the dream of biologists of an earlier generation who anticipated the use of genetic differences among stocks to identify the stock composition of mixed populations (e.g., see Cushing 1956; Ridgway and Klontz 1960). Although the technical Limitations of this era precluded achieving this goal, these pioneering efforts germinated, grew, and ultimately matured as a uniquely valuable tool for fisheries research and management. We regard these new capabilities of the genetic method of stock identification as a complement to rather than a replacement of existing methods of stock identification, such as coded wire tags and scale analyses, which also have unique capabilities, Nevertheless, the genetic method opens up new frontiers for research and management just as these other--and now traditional---methods did upon their introduction. The capability for reliable estimates of component stocks of mixed populations exists wherever known genetic differences exist among contributing stocks. Such differences appear to be the rule rather than the exception as more detailed data accumulate on stocks of anadromous salmonids. This capability is readily transformed into reality once a useable data base is established. It is now time for research and management personnel to become aware of this new capability and to put it to optimal use.

## LITERATURE CITED


Cushing, J. E.
1956. Observations on the serology of tuna. U.S. Fish and Wildlife Service, Special Scientific Report--Fisheries No. 183, 14 pp.

Dempster, A. P., N. M. Laird, and D. B. Rubin.
1977. Maximum likelihood from incomplete data via the EM algorithm. Proc. Royal Statistical Society, Series B Methodol, 39: 1-38.

May, B. P.
1975. Electrophoretic variation in the genus Oncorhynchus: the methodology, genetic basis, and practical applications to fisheries research and management. University of Washington, M.S. thesis.

Milner, G. B., and D. J. Teel.
1979. Columbia River Stock Identification Study: Annual Report of Research (FY 78), NOAA, NWAFC, Seattle, WA 98112.

Milner, G. B., D. J. Teel, and F. M. Utter.
1980. Columbia River Stock Identification Study: Annual Report of Research (FY 79), NOM, NWAFC, Seattle, WA 98112.

Nei, M.
1972. genetic distance between populations. Amer. Natur. 106:283-292.

Ridgway, G. J., and G. W. Klontz.
1960. Blood types in Pacific salmon U.S. fish and Wildlife Service, Special Scientific Report--fisheries No. 324, 9 pp.

Utter, F. M., H. 0. Hodgins, and F. W. Allendorf.
1974. Biochemical genetic studies of fishes: potentialities and limitations. In: Biochemical and Biophysical Perspectives in Marine Biology, Vol. 1, Academic Press, London, New York.

# APPENDIX A

## ALLELIC FREQUENCIES OF TEN POLYMORPHIC LOCI OF SPRING CHINOOK SALMON FROM 14 HATCHERY STOCKS

(Sample sizes refer to number of alleles)

Appendix Table Al.--Protein: Adenosine deaminase Locus: ADA

| Stock | Sample size | Allele frequencies | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Warm Springs | 390 | 0.98 | 0.02 | 0.00 | 0.00 |
| Rapid River | 284 | 0.99 | 0.01 | 0.00 | 0.00 |
| Kooskia | 384 | 0.98 | 0.02 | 0.00 | 0.00 |
| Round Butte | 358 | 0.99 | 0.01 | 0.00 | 0.00 |
| Carson | 392 | 0.98 | 0.02 | 0.00 | 0.00 |
| Eagle Creek | 396 | 1.00 | 0.00 | 0.00 | 0.00 |
| Little White Salmon | 400 | 0.96 | 0.04 | 0.00 | 0.00 |
| South Santiam | 298 | 1.00 | 0.00 | 0.00 | 0.00 |
| Oakridge | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Kalama | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Cowlitz | 378 | 0.97 | 0.03 | 0.00 | 0.00 |
| McKenzie | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Leavenworth | 400 | 0.95 | 0.05 | 0.00 | 0.00 |
| Klickitat | 400 | 0.96 | 0.04 | 0.00 | 0.00 |

Appendix Table A2.--Protein: Alcohol dehydrogenase Locus: ADH

| Stock | Sample size | Allele frequencies | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Warm Springs | 394 | 1.00 | 0.00 | 0.00 | 0.00 |
| Rapid River | 390 | 1.00 | 0.00 | 0.00 | 0.00 |
| Kooskia | 390 | 0.99 | 0.01 | 0.00 | 0.00 |
| Round Butte | 386 | 0.99 | 0.01 | 0.00 | 0.00 |
| Carson | 400 | 0.99 | 0.01 | 0.00 | 0.00 |
| Eagle Creek | 394 | 1.00 | 0.00 | 0.00 | 0.00 |
| Little White Salmon | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| South Santiam | 398 | 1.00 | 0.00 | 0.00 | 0.00 |
| Oakridge | 400 | 0.99 | 0.01 | 0.00 | 0.00 |
| Kalama | 390 | 0.98 | 0.02 | 0.00 | 0.00 |
| Cowlitz | 400 | 0.96 | 0.04 | 0.00 | 0.00 |
| McKenzie | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Leavenworth | 388 | 0.96 | 0.04 | 0.00 | 0.00 |
| Klickitat | 400 | 0.97 | 0.03 | 0.00 | 0.00 |

Appendix Table A3.--Protein: Glycyl-leucine dipeptidase Locus: GL-1

| Stock | Sample size | Allele frequencies | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Warm Springs | 400 | 0.99 | 0.01 | 0.00 | 0.00 |
| Rapid River | 392 | 1.00 | 0.00 | 0.00 | 0.00 |
| Kooskia | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Round Butte | 372 | 1.00 | 0.00 | 0.00 | 0.00 |
| Carson | 396 | 0.99 | 0.01 | 0.00 | 0.00 |
| Eagle Creek | 396 | 1.00 | 0.00 | 0.00 | 0.00 |
| Little White Salmon | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| South Santiam | 394 | 1.00 | 0.00 | 0.00 | 0.00 |
| Oakridge | 398 | 1.00 | 0.00 | 0.00 | 0.00 |
| Kalama | 398 | 0.97 | 0.03 | 0.00 | 0.00 |
| Cowlitz | 396 | 0.98 | 0.02 | 0.00 | 0.00 |
| McKenzie | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Leavenworth | 398 | 0.97 | 0.02 | 0.00 | 0.01 |
| Klickitat | 400 | 1.00 | 0.00 | 0.00 | 0.00 |

Appendix Table A4.--Protein: isocitrate dehydrogenase Locus : IDH

| Stock | Sample size | Allele frequencies 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Warm Springs | 764 | 0.83 | 0.00 | 0.17 | 0.00 |
| Rapid River | 784 | 0.97 | 0.00 | 0.03 | 0.00 |
| Kooskia | 800 | 0.90 | 0.01 | 0.09 | 0.00 |
| Round Butte | 764 | 1 l 00 | 0.00 | 0.00 | 0.00 |
| Carson | 800 | 0.89 | 0.00 | 0.11 | 0.00 |
| EagLe Creek | 788 | 1.00 | 0.00 | 0.00 | 0.00 |
| Little White Salmon | 796 | 0.93 | 0.00 | 0.07 | 0.00 |
| South Santiam | 788 | 0.92 | 0.08 | 0.00 | 0.00 |
| Oakridge | 796 | 0.85 | 0.13 | 0.02 | 0.00 |
| Kalama | 788 | 0.91 | 0.09 | 0.00 | 0.00 |
| Cowlitz | 796 | 0.95 | 0.03 | 0.02 | 0.00 |
| McKenzie | 800 | 0. 91 | 0.09 | 0.00 | 0.00 |
| Leavenworth | 788 | 0.92 | 0.00 | 0.08 | 0.00 |
| Klickitat | 800 | 0.94 | 0.02 | 0.04 | 0.00 |

Appendix Table A5.--Protein: Lactate dehydrogenase Locus: LDH-4

| Stock | Sample size | Allele frequencies | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Warm Springs | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Rapid River | 392 | 0.99 | 0.01 | 0.00 | 0.00 |
| Kooskia | 396 | 0.98 | 0.02 | 0.00 | 0.00 |
| Round Butte | 384 | 1.00 | 0.00 | 0.00 | 0.00 |
| Carson | 398 | 0.98 | 0.02 | 0.00 | 0.00 |
| Eagle Creek | 396 | 1.00 | 0.00 | 0.00 | 0.00 |
| Little White Salmon | 302 | 1.00 | 0.00 | 0.00 | 0.00 |
| South Santiam | 390 | 1.00 | 0.00 | 0.00 | 0.00 |
| Oakridge | 386 | 1.00 | 0.00 | 0.00 | 0.00 |
| Kalama | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Cowlitz | 398 | 1.00 | 0.00 | 0.00 | 0.00 |
| McKenzie | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Leavenworth | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Klickitat | 400 | 1.00 | 0.00 | 0.00 | 0.00 |

Appendix Table A6.--Protein: Lactate dehydrogenase Locus: LDH-5

| Stock | Sample size | Allele frequencies | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Warm Springs | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Rapid River | 392 | 1.00 | 0.00 | 0.00 | 0.00 |
| Kooskia | 396 | 1.00 | 0.00 | 0.00 | 0.00 |
| Round Butte | 382 | 0.98 | 0.02 | 0.00 | 0.00 |
| Carson | 398 | 1.00 | 0.00 | 0.00 | 0.00 |
| Eagle Creek | 396 | 1.00 | 0.00 | 0.00 | 0.00 |
| Little White Salmon | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| South Santiam | 390 | 1.00 | 0.00 | 0.00 | 0.00 |
| Oakridge | 386 | 1.00 | 0.00 | 0.00 | 0.00 |
| Kalama | 398 | 1.00 | 0.00 | 0.00 | 0.00 |
| Cowlitz | 396 | 0.98 | 0.02 | 0.00 | 0.00 |
| McKenzie | 400 | 1.00 | 0.00 | 0.00 | 0.00 |
| Leavenworth | 400 | 0.99 | 0.01 | 0.00 | 0.00 |
| Klickitat | 398 | 0.99 | 0.01 | 0.00 | 0.00 |

Appendix Table A7.--Protein: Leucyl-glycyl-glycine tripeptidase Locus: LGG

| Stock | Sample size | Allele frequencies 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Warm Springs | 398 | 0.94 | 0.06 | 0.00 | 0.00 |
| Rapid River | 378 | 0.90 | 0.10 | 0.00 | 0.00 |
| Kooskia | 396 | 0.93 | 0.07 | 0.00 | 0.00 |
| Round Butte | 380 | 0.96 | 0.04 | 0.00 | 0.00 |
| Carson | 380 | 0.96 | 0.04 | 0.00 | 0.00 |
| Eagle Creek | 386 | 1.00 | 0.00 | 0.00 | 0.00 |
| Little White Salmon | 384 | 0.95 | 0.05 | 0.00 | 0.00 |
| South Santiam | 394 | 0.84 | 0.16 | 0.00 | 0.00 |
| Oakridge | 366 | 0.85 | 0.15 | 0.00 | 0.00 |
| Kalama | 400 | 0.97 | 0.03 | 0.00 | 0.00 |
| Cowlitz | 396 | 0.91 | 0.09 | 0.00 | 0.00 |
| McKenzie | 384 | 0.90 | 0.10 | 0.00 | 0.00 |
| Leavenworth | 396 | 0.97 | 0.03 | 0.00 | 0.00 |
| Klickitat | 396 | 0.93 | 0.07 | 0.00 | 0.00 |

Appendix Table A8.--Protein: Malate dehydrogenase Locus: MDH

| Stock | Sample size | Allele frequencies | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Warm Springs | 800 | 1.00 | 0.00 | 0.00 | 0.00 |
| Rapid River | 772 | 1.00 | 0.00 | 0.00 | 0.00 |
| Kooskia | 788 | 0.97 | 0.03 | 0.00 | 0.00 |
| Round Butte | 764 | 1.00 | 0.00 | 0.00 | 0.00 |
| Carson | 800 | 0.89 | 0.00 | 0.11 | 0.00 |
| Eagle Creek | 792 | 1.00 | 0.00 | 0.00 | 0.00 |
| Little White Salmon | 800 | 0.98 | 0.02 | 0.00 | 0.00 |
| South Santiam | 752 | 0.94 | 0.06 | 0.00 | 0.00 |
| Oakridge | 788 | 0.98 | 0.02 | 0.00 | 0.00 |
| Kalama | 784 | 0.99 | 0.01 | 0.00 | 0.00 |
| Cowlitz | 788 | 0.99 | 0.01 | 0.00 | 0.00 |
| McKenzie | 800 | 0.95 | 0.05 | 0.00 | 0.00 |
| Leavenworth | 792 | 0.99 | 0.01 | 0.00 | 0.00 |
| Klickitat | 800 | 0.98 | 0.02 | 0.00 | 0.00 |

Appendix Table A9.--Protein: Phosphomannose isomerase Locus: PMI

| Stock | Sample size | Allele frequencies 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Warm Springs | 400 | 0.92 | 0.08 | 0.00 | 0.00 |
| Rapid River | 392 | 0.86 | 0.14 | 0.00 | 0.00 |
| Kooskia | 398 | 0.82 | 0.18 | 0.00 | 0.00 |
| Round Butte | 378 | 0.86 | 0.14 | 0.00 | 0.00 |
| Carson | 400 | 0.85 | 0.15 | 0.00 | 0.00 |
| Eagle Creek | 390 | 0.48 | 0.52 | 0.00 | 0.00 |
| Little White Salmon | 400 | 0.82 | 0.18 | 0.00 | 0.00 |
| South Santiam | 396 | 0.55 | 0.45 | 0.00 | 0.00 |
| Oakridge | 400 | 0.43 | 0.57 | 0.00 | 0.00 |
| Kalama | 396 | 0.51 | 0.49 | 0.00 | 0.00 |
| Cowlitz | 398 | 0.49 | 0.50 | 0.01 | 0.00 |
| McKenzie | 400 | 0.45 | 0.55 | 0.00 | 0.00 |
| Leavenworth | 400 | 0.90 | 0.10 | 0.00 | 0.00 |
| Klickitat | 400 | 0.69 | 0.30 | 0.01 | 0.00 |

Appendix Table A10.--Protein: Tetrazolium oxidase Locus: TO

| Stock | Sample size | Allele frequencies 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Warm Springs | 400 | 0.49 | 0.51 | 0.00 | 0.00 |
| Rapid River | 390 | 0.96 | 0.04 | 0.00 | 0.00 |
| Kooskia | 400 | 0.75 | 0.25 | 0.00 | 0.00 |
| Round Butte | 382 | 0.59 | 0.41 | 0.00 | 0.00 |
| Carson | 398 | 0.74 | 0.26 | 0.00 | 0.00 |
| Eagle Creek | 392 | 0.80 | 0.20 | 0.00 | 0.00 |
| Little White Salmon | 394 | 0.78 | 0.22 | 0.00 | 0.00 |
| South Santiam | 396 | 0.85 | 0.15 | 0.00 | 0.00 |
| Oakridge | 400 | 0.92 | 0.08 | 0.00 | 0.00 |
| Kalama | 400 | 0.73 | 0.27 | 0.00 | 0.00 |
| Cowlitz | 400 | 0.61 | 0.39 | 0.00 | 0.00 |
| McKenzie | 400 | 0.86 | 0.14 | 0.00 | 0.00 |
| Leavenworth | 400 | 0.69 | 0.31 | 0.00 | 0.00 |
| Klickitat | 400 | 0.69 | 0.31 | 0.00 | 0.00 |

# APPENDIX B

## ESTIMATED PERCENT CONTRIBUTION, STANDARD DEVIATION OF THE ESTIMATE, AND DIFFERENCE BETWEEN ACTUAL AND ESTIMATED CONTRIBUTION FOR EACH OF FOUR MIXED STOCK SAMPLES HAVING KNOWN COMPOSITIONS

Appendix Table Bl.
Sample A

| Contributor | Contribution | | | Standard deviation |
|---|---|---|---|---|
| | Actual | Estimated | Difference | |
| Stock (14) | | | | |
| Rapid River | 5.7 | 1.5 | -4.2 | 2.1 |
| Kooskia | 5.7 | 9.4 | 3.7 | 10.9 |
| Leavenworth | 11.4 | 9.9 | -1.5 | 4.4 |
| Round Butte | 0.0 | 7.0 | 7.0 | 5.8 |
| Warm Springs | 28.6 | 21.6 | -7.0 | 4.1 |
| Klickitat | 14.3 | 16.8 | 2.5 | 5.5 |
| Little White Salmon | 0.0 | 1.5 | 1.5 | 7.3 |
| Carson | 28.6 | 26.1 | -2.5 | 9.6 |
| Oakridge | 0.0 | 0.0 | 0.0 | 2.0 |
| McKenzie | 0.0 | 0.0 | 0.0 | 3.2 |
| South Santiam | 5.7 | 6.0 | 0.3 | 3.2 |
| Eagle Creek | 0.0 | 0.0 | 0.0 | 1.6 |
| Kalama | 0.0 | 0.0 | 0.0 | 1.8 |
| Cowlitz | 0.0 | 0.0 | 0.0 | 2.7 |
| Geographic group (5) | | | | |
| Snake | 11.4 | 10.9 | -0.5 | 11.0 |
| Upper Columbia | 11.9 | 9.9 | -2.0 | 4.4 |
| Bonneville Pool | 71.5 | 73.0 | 1.5 | 11.8 |
| Willamette | 5.7 | 6.0 | 0.3 | 0.0 |
| Kalama/Cowlitz | 0.0 | 0.0 | 0.0 | 3.2 |
| Geographic group (2) | | | | |
| Above Bonneville | 94.3 | 93.8 | -0.5 | 0.6 |
| Below Bonneville | 5.7 | 6.0 | 0.3 | 0.6 |

Appendix Table B2.
Sample B

| Contributor | Contribution | | | Standard deviation |
|---|---|---|---|---|
| | Actual | Estimated | Difference | |
| Stock (14) | | | | |
| Rapid River | 0.0 | 0.0 | 0.0 | 1.9 |
| Kooskia | 10.0 | 2.9 | -7.1 | 7.5 |
| Leavenworth | 5.0 | 8.6 | 3.6 | 3.7 |
| Round Butte | 0.0 | 6.7 | 6.7 | 4.5 |
| Warm Springs | 5.0 | 0.3 | -4.7 | 2.4 |
| Klickitat | 15.0 | 19.8 | 4.8 | 6.2 |
| Little White Salmon | 0.0 | 1.5 | 1.5 | 4.7 |
| Carson | 0.0 | 0.0 | 0.0 | 5.9 |
| Oakridge | 10.0 | 7.5 | -2.5 | 2.6 |
| McKenzie | 15.0 | 33.4 | 18.4 | 5.8 |
| South Santiam | 10.0 | 0.0 | -10.0 | 4.3 |
| Eagle Creek | 15.0 | 6.2 | -8.8 | 2.4 |
| Kalama | 0.0 | 3.0 | 3.0 | 3.0 |
| Cowlitz | 15.0 | 9.9 | -5.1 | 3.8 |
| Geographic group (5) | | | | |
| Snake | 10.0 | 2.9 | -7.1 | 7.7 |
| Upper Columbia | 5.0 | 8.6 | 3.6 | 3.7 |
| Bonneville Pool | 20.0 | 28.3 | 8.3 | 10.0 |
| Willamette | 50.0 | 47.1 | -2.9 | 0.0 |
| Kalama/Cowlitz | 15.0 | 12.9 | -2.1 | 3.5 |
| Geographic group (2) | | | | |
| Above Bonneville | 35.0 | 39.8 | 4.8 | 1.1 |
| Below Bonneville | 65.0 | 60.0 | -5.0 | 1.1 |

Appendix Table B3.
Sample C

| Contributor | Contribution | | | Standard deviation |
|---|---|---|---|---|
| | Actual | Estimated | Difference | |
| Stock (14) | | | | |
| Rapid River | 30.4 | 30.0 | -0.4 | 2.5 |
| Kooskia | 29.2 | 25.8 | -3.4 | 7.1 |
| Leavenworth | 0.0 | 1.3 | 1.3 | 2.9 |
| Round Butte | 0.0 | 0.0 | 0.0 | 2.8 |
| Warm Springs | 5.8 | 4.4 | -1.4 | 1.8 |
| Klickitat | 0.0 | 4.8 | 4.8 | 3.0 |
| Little White Salmon | 0.0 | 0.0 | 0.0 | 5.7 |
| Carson | 10.3 | 12.5 | 2.2 | 5.7 |
| Oakridge | 12.2 | 6.5 | -5.7 | 1.9 |
| McKenzie | 0.0 | 0.0 | 0.0 | 3.4 |
| South Santiam | 0.0 | 6.4 | 6.4 | 2.9 |
| Eagle Creek | 0.0 | 0.0 | 0.0 | 1.8 |
| Kalama | 0.0 | 0.5 | 0.5 | 1.9 |
| Cowlitz | 12.2 | 7.8 | -4.4 | 2.1 |
| Geographic group (5) | | | | |
| Snake | 59.6 | 55.8 | -3.8 | 7.2 |
| Upper Columbia | 0.0 | 1.3 | 1.3 | 2.9 |
| Bonneville Pool | 16.1 | 21.7 | 5.6 | 8.3 |
| Willamette | 12.2 | 12.9 | 0.7 | 2.4 |
| Kalama/Cowlitz | 12.2 | 8.3 | -3.9 | 2.2 |
| Geographic group (2) | | | | |
| Above Bonneville | 75.7 | 78.8 | 3.1 | 1.0 |
| Below Bonneville | 24.4 | 21.2 | -3.2 | 1.0 |

Appendix Table B4.
Sample D

| Contributor | Contribution | | | Standard deviation |
|---|---|---|---|---|
| | Actual | Estimated | Difference | |
| Stock (14) | | | | |
| Rapid River | 5.8 | 0.0 | -5.8 | 2.3 |
| Kooskia | 12.8 | 17.3 | 4.5 | 11.0 |
| Leavenworth | 12.8 | 4.5 | -8.3 | 5.3 |
| Round Butte | 0.0 | 0.0 | 0.0 | 4.7 |
| Warm Springs | 0.0 | 0.0 | 0.0 | 2.0 |
| Klickitat | 12.8 | 19.3 | 6.5 | 7.1 |
| Little White Salmon | 0.0 | 2.2 | 2.2 | 7.1 |
| Carson | 0.0 | 0.0 | 0.0 | 9.8 |
| Oakridge | 12.8 | 14.3 | 1.5 | 3.3 |
| McKenzie | 5.1 | 16.5 | 11.4 | 7.4 |
| South Santiam | 12.2 | 7.3 | -4.9 | 5.5 |
| Eagle Creek | 12.8 | 6.4 | -6.4 | 2.8 |
| Kalama | 0.0 | 0.0 | 0.0 | 3.5 |
| Cowlitz | 12.8 | 12.1 | -0.7 | 4.0 |
| Geographic group (5) | | | | |
| Snake | 18.6 | 17.3 | -1.3 | 11.4 |
| Upper Columbia | 12.8 | 4.5 | -8.3 | 5.3 |
| Bonneville Pool | 12.8 | 21.5 | 8.7 | 10.0 |
| Willamette | 42.9 | 44.5 | 1.6 | 4.5 |
| Kalama/Cowlitz | 12.8 | 12.1 | -0.7 | 3.2 |
| Geographic group (2) | | | | |
| Above Bonneville | 44.2 | 43.3 | -0.9 | 1.3 |
| Below Bonneville | 55.7 | 56.6 | 0.9 | 1.3 |